

universität ulm¹ **TU WIEN**⁴ **TUM**⁵

Google² **BOSCH**³

RelationField

Relate Anything in Radiance Fields

Sebastian Koch¹ Johanna Wald² Mirco Colosi³ Narunas Vaskevicius³
Pedro Hermosilla⁴ Federico Tombari^{2,5} Timo Ropinski¹

CVPR Nashville JUNE 11-15, 2025

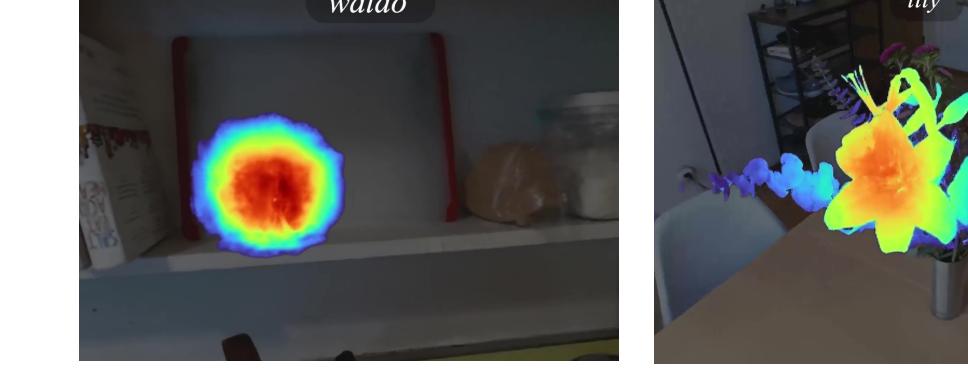
Paper, code, and demo are available: relationfield.github.io/

Contributions

- Interactive querying of relationships in radiance fields
- 3D inter-object relationship reasoning learned from 2D supervision
- Improved open-vocab. object localization with complex queries
- State-of-the-art open-vocabulary 3D Scene Graph prediction

Motivation

Radiance Fields incorporate 3D object semantics learned from 2D views

$x, y, z \rightarrow$  $\rightarrow r, g, b + \sigma$
 $\theta, \varphi \rightarrow$ 



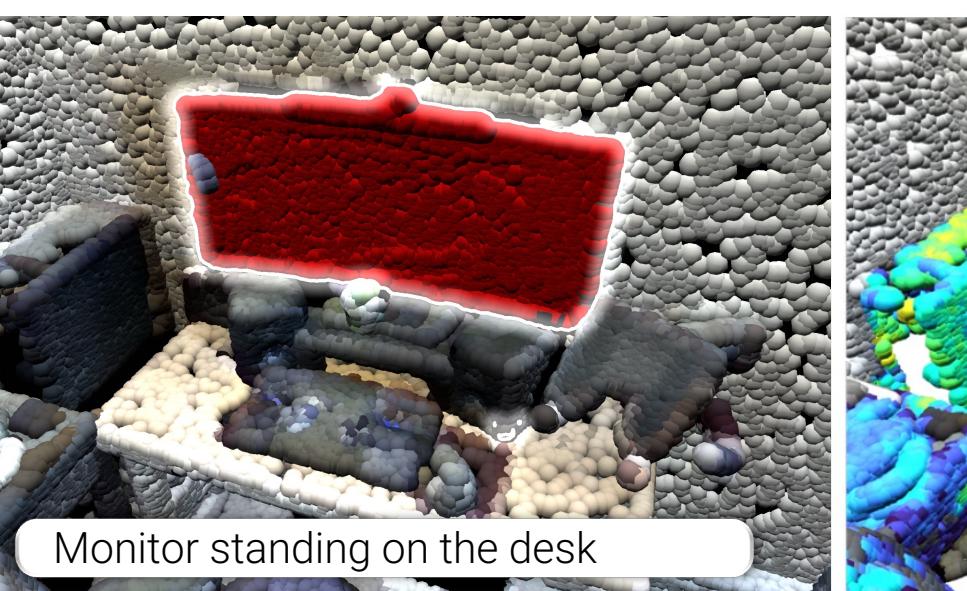
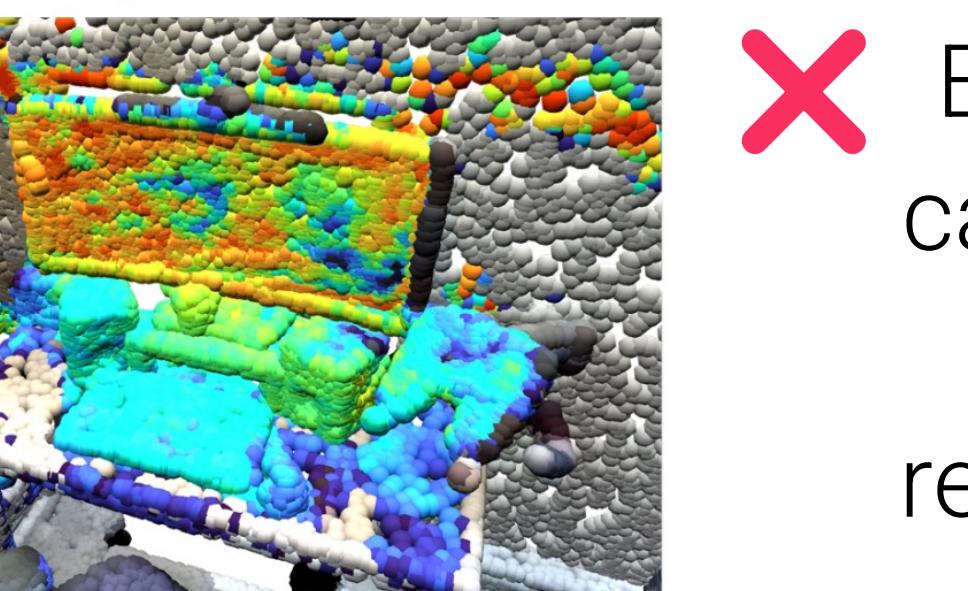
Simple NeRF modification with additional foundation model output

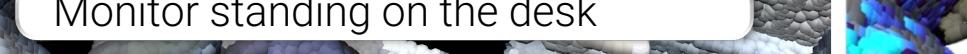
Focus on object-centric representations such as (semantic/instance segmentation)



3D Scene Reasoning requires relationship understanding



Ground Truth 
OpenNeRF 



 Existing methods cannot understand complex relationship queries

TLDR: Query Any Relationship in 3D

Composition, Compare, Spatial, Affordance, Support

RelationField

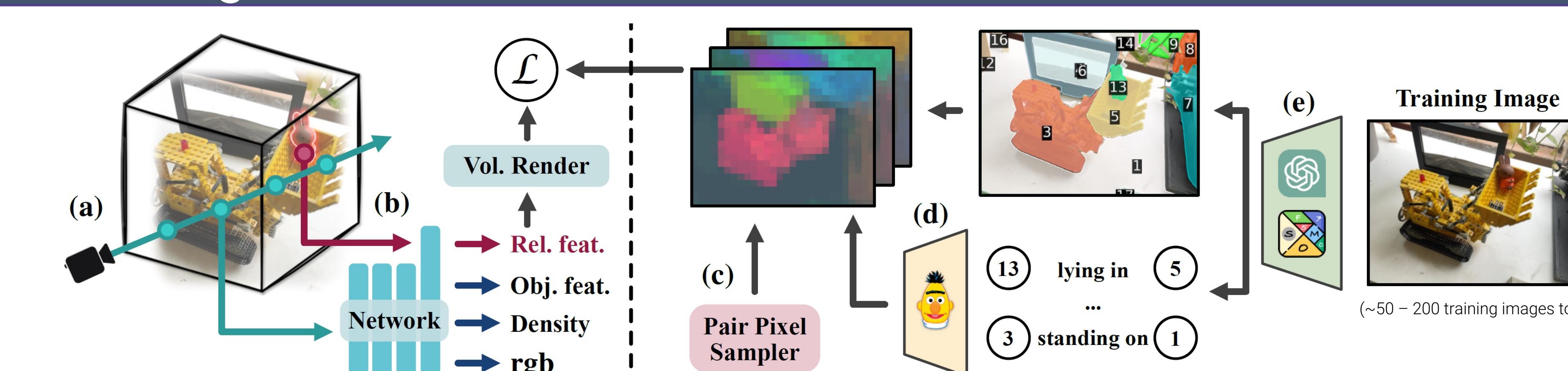
xyz → Network → Rel. feat., Obj. feat., Density, rgb

d → Network → Rel. feat., Obj. feat., Density, rgb

Additional query ray for 3D relationship grounding

Interactive relationship querying

Training RelationField



(a) Training Image (~50 – 200 training images total)
(b) Network (xyz, d) → Rel. feat., Obj. feat., Density, rgb
(c) Pair Pixel Sampler
(d) Vol. Render → Loss (L)
(e) Training Image

Zero-shot open-vocabulary relationships

Bedroom, "standing on", Bathroom, "belongs to", Living room, "standing on", "covers", "turns on", "attached to", "used with", "lying on", "same as"

Experiments

3D Open-Vocabulary Instance Segmentation

Ground Truth, OpenNeRF, RelationField

Curtain hanging on door, Bag-of-words: 'curtain' & 'door', the curtain hanging on the door

Multi-View Consistency

SoM+GPT-4v (2D), Ours (3D)

! 3D awareness is important

Baseline Comparison

IoU: LERF (0.25), OpenNeRF (0.45), LangSplat (0.49), RelationField (0.53). Acc: LERF (0.50), OpenNeRF (0.83), LangSplat (0.87), RelationField (0.96).

! RelationField new SOTA for complex localization queries

Application: 3D Scene Graph Prediction

RelationField enables open-vocab 3D Scene Graph prediction

Point Queries, Relationship Response (lying on), Instance Aggregation (lying on)

A 3D Scene Graph combines object semantics & inter-object relationships

Object R@5 R@10 R@3 R@5
Method GPT-4 [1] (2D+depth) 0.34 0.42 0.55 0.58
Llama 3.2 [12] (2D+depth) 0.40 0.52 0.46 0.48
Open3DSG [27] 0.56 0.61 0.58 0.65
ConceptGraphs [15] 0.37 0.46 0.74 0.79
RelationField **0.69** **0.80** **0.76** **0.82**

SOTA OV 3D Scene Graph prediction